Attorney Docket No. 1668.1021

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Ophir Frieder, et al.

Application No.: 09/629,175   Group Art Unit: 2163

Filed: July 31, 2000   Examiner: Le, Uyen T

For: SYSTEM FOR SIMILAR DOCUMENT DETECTION

REPLY BRIEF UNDER 37 CFR § 41.41

Commissioner for Patents
Board of Patent Appeals and Interferences
United States Patent and Trademark Office
PO Box 1450
Alexandria, VA 22313-1450

Sir:

Pursuant to the Appellant's earlier filed Notice of Appeal on December 9, 2005, the Appellant appealed the Examiner's September 9, 2005 Office Action finally rejecting claims 1-29, 44-48 and 50-57. Appellant's Appeal Brief was filed June 27, 2006. The Examiner's Answer was mailed August 4, 2006. Appellant's Reply Brief together with the requisite fees set forth in 37 CFR § 41.20 is submitted within two months from the mailing date of the Examiner's Answer.

In the present application before the Board of Patent Appeals and Interferences, claims 30-33, 49, and 58-62 have been allowed and claims 1-28, 45-48, and 50-57 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Aiken (U.S. 6,240,409) of record ("the '409 patent").

With respect to the Examiner's rejections, claims 1, 50, and 51 are independent claims. Independent claim 1 includes the following important limitation: "filtering the document to eliminate tokens and obtain a filtered document containing remaining tokens, the tokens being eliminated based on at <u>least one of (a) parts of speech and (b) collection statistics</u> relating to a number of occurrences of words or phrases in the document". Independent method claim 50 discloses, "filtering the document to eliminate tokens based on <u>parts of speech</u> and obtain a filtered document" as an important limitation. Lastly, independent claim 51 includes "a filter to filter the document to eliminate tokens based on <u>parts of speech</u> and obtain a filtered document" as a limitation.

Spanning pages 9-11 of the Examiner's Answer, the Examiner sets forth various issues, which the Examiner believes to be central points in the Appellant's Brief filed on August 19, 2005. Issues 1 and 2 address whether claims 50 and 51 read on the prior art's removal of "stop words." Issue 3 addresses whether the present invention explicitly distinguishes between token removal based on parts of speech and token removal based on stop words. Issue 4 addresses whether the present invention reads on the prior art's removal use of frequently used words. These issues will be addressed in turn.

ISSUE 1

The Examiner addresses Applicants' arguments regarding whether the present invention reads on the '409 patent removal of stop words as follows:

> The claim language does not require filtering based on <u>all</u> parts of speech of a spoken language, thus interpreted broadly, <u>any</u> part of speech for example article, preposition that happens to be a stop word would meet the limitation of filtering based on parts of speech. (Emphasis added)

By interpreting claims 50 and 51 thusly, the Examiner has not given full effect to the verbiage of these claims. For example, the phrase "parts of speech" appears in the following element of claim 50: "filtering the documents to eliminate tokens based on

2

parts of speech and obtain a filtered document." Claim 51 has a similar structure with respect to the phrase "parts of speech." In addressing the Applicants' arguments, the Examiner injects the modifying word "all" and "any" to the structure of claims 50 and 51. The phase "parts of speech," as it appears in claims 50 and 51 is not modified by either the word "all" or "any" and thus the claims do not require, according to the Examiner's Answer, "filtering based on all parts of speech" or "any part of speech." However, the claims do require filtering based on parts of speech. Thus, a first action is taken if a predetermined part of speech occurs, and the first action is not taken if a part of speech *different from* the predetermined part of speech occurs. The Examiner refers to *all* parts of speech and *any* part of speech. However, for there to be filtering based on parts of speech, there must be a part of speech different from a predetermined part of speech. If one attempts to filter based on *any* part of speech or *all* parts of speech, then there is no filtering. That is, a first action is always taken. Filtering requires for it to be possible that sometimes the first action is not taken. Giving proper effect to the verbiage of claims 50 and 51 would provide *uniform* treatment of various parts of speech in contrast to filtering based on any or all parts of speech.

Therefore, the '409 patent, by simply filtering out stop words, does not provide uniform treatment to various parts of speech. There is no uniform treatment of any particular part of speech. For example, some prepositions trigger the method of '409 patent, and some prepositions do not trigger the method of the '409 patent. In order for there to be filtering based on parts of speech, this example requires that all prepositions either trigger the method or do not trigger the method. Otherwise, it is clear that the method is completely unconcerned with whether a word is a preposition or a verb or a noun, etc. Consequently, the '409 patent discloses a different criteria for filtering and thus has a different scope to the language of claims 50 and 51.

ISSUE 2

The '409 patent discloses the following:

1. "Further processing can include removing <u>unimportant</u> words such as 'the' or 'and'." (Emphasis added) (See column 4, lines 57-58 of the '409 patent).

2. "Further preprocessing of raw data strings could include removing words 'this' and 'is' under the assumption that they are words that would be used frequently and would not be useful indicators of copying." (Emphasis added) (See column 8, line 67 - column 9, line 3 of the '409 patent).

The Examiner's Answer responds to the Applicants' arguments as follows:

although some words in a spoken language are considered stop words, all words must be parts of speech . . . the examiner maintains that the stop words and words frequently used in [the '409 patent] read on the parts of speech of claims 1, 50, and 51."

Once again, the Examiner has not given proper effect to the wording of the claims. The Applicants' agree that "all words must be parts of speech," however, claims 1, 50, and 51 do not filter based on any part of speech. Instead, these claims give uniform treatment to at least one particular part of speech. Moreover, the '409 patent teaches away from such uniform treatment of various parts of speech by explicitly limiting itself to both "unimportant" words and words "used frequently."

There are eight commonly accepted parts of speech, namely:

1. Nouns

2. Verbs

3. Adjectives

4. Adverbs

5. Pronouns

6. Prepositions

7. Conjunctions

8. Interjections

Thus, even if one equates articles to stop words, articles are **not** commonly accepted as a "part of speech". However, even considering articles as a "part of speech," which they are not, in the rejection, the Examiner implies that all articles are included in stop word lists. Hence, the Examiner claims that by removing stop words, one is effectively filtering by a part of speech, namely "articles". However, it is not

necessarily the case that all articles are included in a stop word list. This is because if one eliminates stop words, particularly when it comes to searching (which is why stop words are used), at times, the results are negatively impacted. For example, "the" (an article) is not always included in a stop-word list since its removal can significantly alter the search. In a **Google**™ search for "the online" the top entry, and the entry of choice is the "Technological Horizons in Education (THE) Journal". THE Online™ is a leading Technology based education publication for K – 12 and higher education. On the other hand, in a **Google**™ search for "online" (removing "the"), THE Online™ is **not** found in any of the first ten screens.

Besides the above, there are at least two major problems with the Examiner's interpretation of the '409 patent:

1. Besides the above inappropriate use of "stop words" as parts of speech, the examiner misquotes '409 patent's stated intention:

   - Regarding the assertion by the examiner of "stop word removal" in column 4 lines 57-58: the '409 patent actually describes the removal of "unimportant words". Clearly, nouns, verbs, etc are not unimportant words. Thus, filtration based on "unimportant words" clearly does not teach our invention.

   - Regarding the assertion by the examiner of "stop word removal" in column 8 line 67 – column 9 line 3: the '409 patent actually describes the removal of "words that would be used frequently anyway and would not be useful indicators ". Clearly, nouns, verbs, etc are useful indicators. Thus, filtration based on "not useful indicators" clearly does not teach our invention.

   - Furthermore, in the same cited notations, the '409 patent exemplifies "stop words" as: "this", "is", "the", "and". Strictly speaking, three of these terms are not articles and not all articles are included in the list. Hence, the '409 patent does not even teach to filter all articles.

2. The examiner states:

"Although [the '409 patent] does not specifically show sorting the filtered document to reorder the tokens according to a predetermined ranking, official notice is taken that it is well known in the art that different operating systems use different tokens ordering. Therefore, it would have been obvious to one of ordinary skill in the art to include sorting the filtered document to reorder the tokens according to a predetermined ranking in order to accommodate different operating systems while implementing the method of [the '409 patent]."

- First, applicants question and request documents showing different operating systems requiring reordering. Applicants do not understand how different operating systems relate to the '409 patent.

- Furthermore, even if one does wish to "accommodate different operating systems", **the '409 patent specifically teaches to preserve order and not sort.** The '409 patent states in column 4 lines 45-47: "the string is translated to a token string that represents <u>and preserves the structure</u> and content of the original or raw data string." (emphasis added)

ISSUE 3

The Examiner states that "Claim 8 . . . clearly shows no distinction between a stop word and a part of speech." The Applicants are baffled by this assertion. Claim 8 depends on claim 2, which in turn depends on claim 1. Dependent claim 8 recites "wherein the step of filtering <u>further</u> comprises removing from the token stream, at least one token corresponding to a stop word". It appears that the Examiner has ignored the term "further," and is again misinterpreting the claims by not considering how the terms relate to one another. Claim 8 <u>does not</u> describe that filtering based on parts of speech is filtering based on stop words. If anything, the principle of claim differentiation, when applied to claim 8, shows that filtering based on stop words is not equivalent to filtering based on parts of speech.

As the language of claim 8 indicates, when the filtering occurs, there is filtering based on parts of speech and there is filtering based on stop words. The two are separate. Thus, the present invention explicitly distinguishes token removal based on parts of speech and token removal based on stop words.

ISSUE 4

Independent claim 1 recites filtering the document to eliminate tokens based on at least one (a) parts of speech and (b) collection statistics relating to a number of occurrences of words or phrases. The other independent claims do not refer to "collection statistics." In the Examiner's Answer, the Examiner states "the number of occurrences of frequent words in [the '409 patent] reads on the claimed collection statistics relating to a number of occurrences of words or phrases in the document." The Applicants do not believe the Examiner is considering the modifier "collection" to the word "statistics," as it appears in claim 1. The Applicants submit that the Examiner is considering usage statistics of an entire language when concluding that "frequent words", without modification or reference to what forms the basis of the frequency of these words, reads on claim 1. Interpreting the language of claim 1 to read on frequent words that appear in any given language ignores the limitation "collection".

The phrase "document collection" is disclosed in the specification to refer to "document collection-specific stop words" (page 12, line 5). Page 4, lines 14-22 describe that similar documents can populate a document collection when multiple document sources are used. This excerpt describes that the National Center for Complimentary and Alternative Medicine supports an information search and retrieval engine for a document collection of medical data having inputs from multiple sources of medical data. Because this excerpt refers to multiple sources of medical data, it is clear that the invention does not relate to the entire English language or all documents. According to the example, the members of the collection are documents from the multiple sources of medical data.

As discussed above, the operative word is "collection," which does not refer to frequently used words in a given language. For example, if the "collection" is patent cases heard by the Board, the phrase "35 U.S.C." would be a frequently occurring phrase. The phrase "35 U.S.C." is not very common in the English language. Therefore, if the '409 patent wanted to eliminate frequently occurring phrases, the phrase "35 U.S.C." would not be eliminated. On the otherhand, if one wanted to eliminate frequently occurring phrases from a collection of patent cases heard by the Board, then "35 U.S.C." may be eliminated. Clearly, the '409 patent uses a different method than the present invention. Thus, the present application gives meaning to the modifier "collection" as it appears in the claims and descriptions of the embodiments.

In addition, the Examiner does not appear to be following current Court of Appeals for the Federal Circuit ("CAFC") jurisprudence with respect to interpreting the language of claim 1. Particularly, the Examiner's Answer states on page 11, lines 8 and 9 "claim 1 does not require both limitations (a) and (b), only at least one of the two." This argument addresses the language "eliminated based on at least one of (a)...and (b)" appearing in claim 1, lines 5-6. This interpretation of claim 1 is in direct contradiction to the CAFC opinion Superguide Corp. v. DirecTV Enterprises, Inc., 69 USPQ2d 1865 (Fed. Cir. 2004). In Superguide, the CAFC upheld the district court's interpretation that "at least one of" followed by a conjunctive list requires selection of at least one value from each member of the list. Id. at 1878. Thus, claim 1 requires filtering based on both parts of speech and collection statistics. As discussed above, the '409 patent does not teach or suggest filtering based on either parts of speech or collection statistics, and the '409 patent does not teach or suggest filtering based on both parts of speech and collection statistics.

CONCLUSION

In view of the law and facts stated herein, the Appellants respectfully maintains that the reasoning and the references cited by the Examiner are insufficient to maintain either an anticipation rejection or an obviousness rejection of the claims. Appellant respectfully urges that the rejections are improper. Reversal of the rejections in this appeal is respectfully requested.
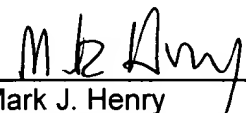
The Commissioner is hereby authorized to charge any additional fees required in connection with the filing of the Appeal Brief to our Deposit Account No. 19-3935.

Respectfully submitted,

STAAS & HALSEY LLP

Date: _October 4, 2006_   By: _M̲ ̲J̲z̲ ̲H̲e̲n̲r̲y̲_

Mark J. Henry
Registration No. 36,162

1201 New York Ave, N.W., Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501